

ΕΦΑΡΜΟΣΜΕΝΗ ΙΑΤΡΙΚΗ ΕΡΕΥΝΑ APPLIED MEDICAL RESEARCH

Εγκυρότητα και αξιοπιστία των ερωτηματολογίων στις επιδημιολογικές μελέτες

1. Εισαγωγή
2. Εγκυρότητα
 - 2.1. Εγκυρότητα περιεχομένου
 - 2.2. Εγκυρότητα εννοιολογικής κατασκευής
 - 2.2.1. Παραγοντική ανάλυση
 - 2.2.2. Συγκλίνοσα εγκυρότητα
 - 2.2.3. Διακρίνοσα εγκυρότητα
 - 2.2.4. Μέθοδος των γνωστών ομάδων
 - 2.3. Εγκυρότητα κριτηρίου
 - 2.3.1. Ταυτόχρονη εγκυρότητα
 - 2.3.2. Προβλεπτική εγκυρότητα
 - 2.4. Εγκυρότητα όψης
3. Αξιοπιστία
 - 3.1. Αξιοπιστία ελέγχου-επανελέγχου
 - 3.2. Αξιοπιστία εναλλακτικών μορφών
 - 3.3. Αξιοπιστία ημίσεων τμημάτων
 - 3.4. Αξιοπιστία εσωτερικής συνέπειας
 - 3.5. Αξιοπιστία μεταξύ παρατηρητών
 - 3.5.1. Δείκτης κάππα
 - 3.5.2. Συντελεστής συσχέτισης εντός ομάδων
4. Επί μέρους ενότιτες ενός ερωτηματολογίου
5. Σύνοψη

1. ΕΙΣΑΓΩΓΗ

Στις επιδημιολογικές μελέτες που διεξάγονται στην εφαρμοσμένη έρευνα στις επιστήμες υγείας χρησιμοποιούνται ερωτηματολόγια για τη συλλογή των απαραίτητων δεδομένων. Ακολούθως, τα δεδομένα αυτά αναλύονται και εξάγονται τα συμπεράσματα των μελετών. Είναι σαφές ότι η εγκυρότητα και η αξιοπιστία των συμπερασμάτων των μελετών καθορίζονται σε σημαντικό βαθμό και από την καταλληλότητα των ερωτηματολογίων που χρησιμοποιούνται από τους ερευνητές.¹⁻⁵ Πιο συγκεκριμένα, τα ερωτηματολόγια που χρησιμοποιούνται στις επιδημιολογικές μελέτες είναι απαραίτητο να χαρακτηρίζονται από υψηλή εγκυρότητα (validity) και αξιοπιστία (reliability). Επισημαίνεται ότι η εγκυρότητα και η αξιοπιστία ενός ερωτηματολογίου δεν αποτελούν εγγενή χαρακτηριστικά του ερωτηματολογίου, εάν δηλαδή ένα ερωτηματολόγιο εμφανίσει υψηλή εγκυρότητα και αξιοπιστία σε ένα μελετώμενο πληθυσμό δεν

σημαίνει απαραίτητα ότι θα εμφανίσει υψηλή εγκυρότητα και αξιοπιστία και σε έναν άλλο μελετώμενο πληθυσμό. Για το λόγο αυτόν, σε κάθε μελέτη είναι απαραίτητο να ελέγχεται η εγκυρότητα και η αξιοπιστία του ερωτηματολογίου που χρησιμοποιείται.

Δυστυχώς, όμως, αρκετοί επιστήμονες υγείας δεν έχουν αντιληφθεί την αναγκαιότητα χρήσης έγκυρων και αξιόπιστων ερωτηματολογίων στις επιδημιολογικές μελέτες, με αποτέλεσμα σε αρκετές περιπτώσεις να μη δίνουν ιδιαίτερη σημασία στα ερωτηματολόγια που χρησιμοποιούνται. Στην περίπτωση αυτή, το πρόβλημα καθίσταται σαφές, αλλά, δυστυχώς, μη αναστρέψιμο τόσο στην ανάλυση των δεδομένων όσο και στη συζήτηση των αποτελεσμάτων της μελέτης.

Λάθη και παραλείψεις που γίνονται αντιληπτά από τους ερευνητές έπειτα από τη συμπλήρωση των ερωτηματολογίων από τους συμμετέχοντες και την ολοκλήρω-

ΑΡΧΕΙΑ ΕΛΛΗΝΙΚΗΣ ΙΑΤΡΙΚΗΣ 2013, 30(1):97-110
ARCHIVES OF HELLENIC MEDICINE 2013, 30(1):97-110

Π. Γαλάνης

Εργαστήριο Οργάνωσης και
Αξιολόγησης Υπηρεσιών Υγείας,
Τμήμα Νοσηλευτικής, Εθνικό και
Καποδιστριακό Πανεπιστήμιο Αθηνών,
Αθήνα

Validity and reliability of
questionnaires in epidemiological
studies

Abstract at the end of the article

Λέξεις ευρητηρίου

Αξιοπιστία
Εγκυρότητα
Ερωτηματολόγιο
Παραγοντική ανάλυση

Υποβλήθηκε 31.8.2012
Εγκρίθηκε 10.9.2012

ση της μελέτης δεν είναι δυνατόν να διορθωθούν, παρά μόνο εάν διεξαχθεί εκ νέου η μελέτη με τις απαιτούμενες διορθώσεις στο ερωτηματολόγιο, γεγονός όμως που είναι πρακτικά ανέφικτο, καθώς αυξάνεται σημαντικά ο χρόνος και το κόστος διεξαγωγής της μελέτης. Για το λόγο αυτόν, είναι απαραίτητο να διεξάγεται μια πιλοτική μελέτη (pilot study) με σχετικά μικρό αριθμό συμμετεχόντων (περίπου 30–50), έτσι ώστε να πραγματοποιείται μια προκαταρκτική εκτίμηση της εγκυρότητας και της αξιοπιστίας του ερωτηματολογίου και να διορθώνονται όσο το δυνατόν περισσότερο τα λάθη και οι παραλείψεις. Ακολούθως, το διορθωμένο ερωτηματολόγιο υποβάλλεται εκ νέου σε έλεγχο της εγκυρότητας και της αξιοπιστίας του, μέσω μιας νέας πιλοτικής μελέτης. Εάν το ερωτηματολόγιο εμφανίσει αποδεκτή εγκυρότητα και αξιοπιστία, τότε αρχίζει πλέον να διανέμεται στους υποψήφιους συμμετέχοντες. Με τον τρόπο αυτόν αρχίζει ουσιαστικά η διεξαγωγή της μελέτης, οπότε και συλλέγονται τα απαιτούμενα δεδομένα με τη χρήση του ερωτηματολογίου. Σε πρόσφατο άρθρο⁶ αναλύθηκαν οι βασικές αρχές διαμόρφωσης και χρήσης των κατάλληλων ερωτηματολογίων στις επιδημιολογικές μελέτες, ενώ στο παρόν άρθρο αναλύονται οι δύο βασικές ιδιότητες που πρέπει να έχει ένα ερωτηματολόγιο και οι οποίες είναι η εγκυρότητα και η ακρίβεια.

2. ΕΓΚΥΡΟΤΗΤΑ

Ο όρος «εγκυρότητα» προέρχεται από τη λατινική λέξη “validus”, που σημαίνει ισχυρός και έχει διάφορες ερμηνείες ανάλογα με το ουσιαστικό ή το επίθετο με το οποίο συνοδεύεται. Στο σημείο αυτό επισημαίνεται η διαφορά ανάμεσα στην εγκυρότητα ενός ερωτηματολογίου και στην εσωτερική εγκυρότητα (internal validity) και την εξωτερική εγκυρότητα (external validity) μιας μελέτης.⁷ Εάν σε μια μελέτη η διαφορά στην έκβαση μεταξύ εκτεθειμένων και μη εκτεθειμένων οφείλεται αποκλειστικά στην έκθεση και όχι στην ύπαρξη συστηματικών σφαλμάτων, τότε η μελέτη αυτή χαρακτηρίζεται από υψηλή εσωτερική εγκυρότητα. Για παράδειγμα, εάν σε μια κλινική δοκιμή για την εκτίμηση της αποτελεσματικότητας μιας νέας φαρμακευτικής αγωγής έναντι της καθιερωμένης το ποσοστό πενταετούς επιβίωσης με τη νέα αγωγή είναι 40% και με την καθιερωμένη αγωγή είναι 30% και η διαφορά αυτή (=40–30%=10%) στα ποσοστά επιβίωσης θεωρηθεί ότι οφείλεται αποκλειστικά στη δράση της νέας αγωγής και όχι στην ύπαρξη συστηματικών σφαλμάτων, τότε η μελέτη χαρακτηρίζεται από υψηλή εσωτερική εγκυρότητα. Στην πράξη, η μείωση του συστηματικού σφάλματος (συστηματικό σφάλμα επιλογής, συστηματικό σφάλμα πληροφορίας και συγχυτές) σε μια μελέτη συνεπάγεται αύξηση της εσωτερικής εγκυρότητας.^{8,9}

Η εξωτερική εγκυρότητα μιας μελέτης, εξ άλλου, αφορά στο βαθμό στον οποίο τα αποτελέσματα της μελέτης, στην οποία συμμετέχουν συγκεκριμένα άτομα, αποτελώντας το μελετώμενο πληθυσμό, μπορούν να γενικευτούν σε έναν ευρύτερο πληθυσμό. Για παράδειγμα, τα αποτελέσματα μιας μελέτης στην οποία συμμετέχουν άνδρες ηλικίας 20–30 ετών δεν μπορούν να γενικευτούν σε γυναίκες ή σε άνδρες ηλικίας >60 ετών. Σημειώνεται ότι η εκτίμηση της εξωτερικής εγκυρότητας είναι περισσότερο υποκειμενική σε σχέση με την εκτίμηση της εσωτερικής εγκυρότητας.

Η εγκυρότητα ενός ερωτηματολογίου αφορά στο βαθμό στον οποίο το ερωτηματολόγιο μετρά την έννοια ή, αλλιώς, τη μεταβλητή την οποία διατείνεται ότι μετρά.^{10–23} Η αύξηση της εγκυρότητας ενός ερωτηματολογίου συνεπάγεται τη μείωση του συστηματικού σφάλματος (systematic error). Το συστηματικό σφάλμα είναι η διαφορά μεταξύ της πραγματικής τιμής και της τιμής που μετράται εμπειρικά.^{8,9} Στην περίπτωση του συστηματικού σφάλματος απαιτείται είτε η γνώση της πραγματικής τιμής είτε η ύπαρξη μιας μεθόδου (μέθοδος αναφοράς ή χρυσός κανόνας*) που θεωρείται ότι μετρά την πραγματική τιμή.** Πρόκειται ουσιαστικά για την επίδιξη μιας ιδανικής κατάστασης, η οποία όμως είναι αδύνατον να επιτευχθεί πρακτικά, καθώς δεν είναι γνωστή η πραγματική τιμή της έννοιας που μετράται με ένα ερωτηματολόγιο. Για το λόγο αυτόν, είναι πρακτικά ανέφικτη η δημιουργία ενός ερωτηματολογίου με τέλεια εγκυρότητα. Εν τούτοις, οι ερευνητές συνεχώς προσπαθούν να βελτιώσουν την εγκυρότητα των ερωτηματολογίων, αποβλέποντας στο να μετρήσουν όσο το δυνατόν πιο έγκυρα τις διάφορες έννοιες.

Η εκτίμηση της εγκυρότητας ενός ερωτηματολογίου περιλαμβάνει την εκτίμηση της εγκυρότητας περιεχομένου, της εγκυρότητας εννοιολογικής κατασκευής, της εγκυρότητας κριτηρίου και της εγκυρότητας όψης.^{10–23}

2.1. Εγκυρότητα περιεχομένου

Η εγκυρότητα περιεχομένου (content validity) αφορά στον τρόπο με τον οποίο κατασκευάζεται ένα ερωτηματο-

* Μέθοδος αναφοράς ή, αλλιώς, χρυσός κανόνας (gold standard) είναι ο όρος που χρησιμοποιείται για να περιγράψει μια μέθοδο, μια διαδικασία ή μια μέτρηση, η οποία μπορεί (ή θεωρείται κατόπιν συμφωνίας ότι μπορεί) να μετρά την πραγματική τιμή. Συχνά, χρησιμοποιείται για τη σύγκριση με νέες μεθόδους.^{8,9}

** Για παράδειγμα, εάν το πραγματικό βάρος ενός ατόμου (το οποίο μετράται με μια ζυγαριά που θεωρείται ως μέθοδος αναφοράς) είναι 80 kg και μια ζυγαριά υπολογίσει το βάρος του ίσο με 79 kg, τότε η ζυγαριά αυτή χαρακτηρίζεται από συστηματικό σφάλμα και μειωμένη εγκυρότητα. Είναι σαφές ότι η δυσκολία έγκειται στο να υπάρξει μια ζυγαριά που να μετρά το πραγματικό βάρος, μια ζυγαριά δηλαδή που να μην έχει συστηματικό σφάλμα.

λόγιο και είναι ο βαθμός στον οποίο ένα ερωτηματολόγιο καλύπτει τις διαφορετικές διαστάσεις της έννοιας ή, αλλιώς, της μεταβλητής που μετράται.^{3,17-24} Για παράδειγμα, εάν υποθεθεί ότι η ποιότητα ζωής περιλαμβάνει έξι διαστάσεις, και πιο συγκεκριμένα τη γενική υγεία, τη σωματική λειτουργικότητα, τη ζωτικότητα, την κοινωνική λειτουργικότητα, το συναισθηματικό ρόλο και την ψυχική υγεία, και σε μια μελέτη για την εκτίμηση της ποιότητας ζωής χρησιμοποιηθεί ένα ερωτηματολόγιο με τέσσερις διαστάσεις (γενική υγεία, σωματική λειτουργικότητα, ζωτικότητα και κοινωνική λειτουργικότητα), τότε είναι σαφές ότι το ερωτηματολόγιο αυτό έχει μειωμένη εγκυρότητα περιεχομένου, καθώς δεν μετρώνται όλες οι διαστάσεις της ποιότητας ζωής.

Τα στοιχεία (ερωτήσεις ή δηλώσεις/προτάσεις) που μπορούν να χρησιμοποιηθούν σε ένα ερωτηματολόγιο για τη μέτρηση μιας συγκεκριμένης έννοιας, θεωρητικά είναι άπειρα. Για παράδειγμα, για τη μέτρηση της ποιότητας ζωής αναφορικά με την ψυχική υγεία, τα στοιχεία που μπορούν να συμπεριληφθούν σε ένα ερωτηματολόγιο είναι άπειρα. Έτσι, για την εκτίμηση της εγκυρότητας περιεχομένου ενός ερωτηματολογίου, το ερώτημα που τίθεται είναι εάν τα στοιχεία του ερωτηματολογίου είναι αντιπροσωπευτικά της ευρύτερης και αφηρημένης έννοιας που πρόκειται να μετρηθεί με το συγκεκριμένο ερωτηματολόγιο. Για παράδειγμα, εάν για τη μέτρηση της ποιότητας ζωής αναφορικά με την ψυχική υγεία χρησιμοποιηθεί ένα ερωτηματολόγιο με 15 στοιχεία, το ερώτημα που τίθεται είναι εάν τα στοιχεία αυτά είναι αντιπροσωπευτικά για την έγκυρη μέτρηση της συγκεκριμένης έννοιας.

Η εγκυρότητα περιεχομένου, ουσιαστικά, αφορά στα στοιχεία που επιλέγονται να συμπεριληφθούν σε ένα ερωτηματολόγιο για τη μέτρηση μιας έννοιας. Τα στοιχεία αυτά δημιουργούνται από τους ερευνητές με βάση την εμπειρία τους, τις γνώσεις τους και βέβαια την κατάλληλη συστηματική βιβλιογραφική ανασκόπηση αναφορικά με την έννοια που επιθυμούν να μετρήσουν. Για το λόγο αυτόν, κρίνεται σκόπιμο κατά την κατασκευή ενός ερωτηματολογίου να αναφέρονται με σαφήνεια οι βιβλιογραφικές πηγές των ερευνητών, έτσι ώστε να είναι δυνατόν να κριθεί η ποιότητα της βιβλιογραφικής ανασκόπησης που διενεργήθηκε.

Επί πλέον, ο έλεγχος της εγκυρότητας περιεχομένου ενός ερωτηματολογίου είναι απαραίτητο να περιλαμβάνει τον υπολογισμό του λόγου εγκυρότητας περιεχομένου (content validity ratio) για κάθε στοιχείο του ερωτηματολογίου.^{24,25} Στην περίπτωση αυτή, οι ερευνητές που κατασκευάζουν ένα ερωτηματολόγιο δημιουργούν μια λίστα από πιθανά στοιχεία που θεωρούν ότι μπορούν να συμπεριληφθούν στο τελικό ερωτηματολόγιο. Ακολούθως, απευθύνονται σε μια ομάδα ερευνητών που θεωρούνται ειδικοί αναφορικά

με την έννοια που μετρά το ερωτηματολόγιο. Για παράδειγμα, εάν ένα ερωτηματολόγιο μετρά την ικανοποίηση των μεταναστών από τις παρεχόμενες δημόσιες υπηρεσίες υγείας στην Αθήνα, τότε οι ειδικοί στους οποίους θα μπορούσε να απευθυνθεί η ερευνητική ομάδα δημιουργίας του ερωτηματολογίου είναι επαγγελματίες υγείας που εργάζονται στα εξωτερικά ιατρεία και στα τμήματα επειγόντων περιστατικών δημόσιων νοσοκομείων στην Αθήνα, καθώς και ερευνητές με επαρκή αριθμό δημοσιεύσεων σε επιστημονικά περιοδικά αναφορικά με τους μετανάστες. Έπειτα, οι δημιουργοί του ερωτηματολογίου ζητούν από την ομάδα των ειδικών να κρίνουν κάθε στοιχείο του ερωτηματολογίου ως «απαραίτητο», «χρήσιμο, άλλα όχι απαραίτητο» ή «μη αναγκαίο» και ακολούθως υπολογίζεται ο λόγος εγκυρότητας περιεχομένου για κάθε στοιχείο του ερωτηματολογίου, σύμφωνα με την παρακάτω ισότητα:^{24,25}

$$\text{Λόγος εγκυρότητας περιεχομένου} = \frac{n_e - \frac{N}{2}}{\frac{N}{2}} \quad (1)$$

Στην ισότητα 1, με N συμβολίζεται ο συνολικός αριθμός των ειδικών που κρίνουν τα στοιχεία του ερωτηματολογίου και με n_e συμβολίζεται ο αριθμός των ειδικών που χαρακτηρίζουν ένα στοιχείο του ερωτηματολογίου ως «απαραίτητο». Για παράδειγμα, εάν ο συνολικός αριθμός των ειδικών είναι 10 και ένα στοιχείο του ερωτηματολογίου χαρακτηρίστηκε ως «απαραίτητο» και από τους 10 ειδικούς, τότε, σύμφωνα με την ισότητα 1, προκύπτει ότι ο λόγος εγκυρότητας περιεχομένου είναι ίσος με 1. Επί πλέον, εάν ο συνολικός αριθμός των ειδικών είναι 10 και ένα στοιχείο του ερωτηματολογίου χαρακτηρίστηκε ως «απαραίτητο» από τους 5 ειδικούς, τότε, σύμφωνα με την ισότητα 1, προκύπτει ότι ο λόγος εγκυρότητας περιεχομένου είναι ίσος με 0. Εάν, εξ άλλου, ο συνολικός αριθμός των ειδικών είναι 10 και ένα στοιχείο του ερωτηματολογίου δεν χαρακτηρίστηκε ως «απαραίτητο» από κανέναν από τους ειδικούς, τότε, σύμφωνα με την ισότητα 1, προκύπτει ότι ο λόγος εγκυρότητας περιεχομένου είναι ίσος με -1. Ο λόγος εγκυρότητας περιεχομένου για ένα στοιχείο ενός ερωτηματολογίου λαμβάνει τιμές από -1 έως 1, με τις μεγαλύτερες τιμές να δηλώνουν μεγαλύτερη εγκυρότητα περιεχομένου. Όταν ο λόγος εγκυρότητας περιεχομένου για ένα στοιχείο ενός ερωτηματολογίου είναι ίσος με 0, τότε οι μισοί ειδικοί θεωρούν το στοιχείο αυτό ως «απαραίτητο». Όταν ο λόγος εγκυρότητας περιεχομένου για ένα στοιχείο ενός ερωτηματολογίου είναι >0, τότε περισσότεροι από τους μισούς ειδικούς θεωρούν το στοιχείο αυτό ως «απαραίτητο». Όταν ο λόγος εγκυρότητας περιεχομένου για ένα στοιχείο ενός ερωτηματολογίου είναι <0, τότε λιγότεροι από τους

μισούς ειδικούς θεωρούν το στοιχείο αυτό ως «απαραίτητο». Είναι σαφές ότι τα στοιχεία ενός ερωτηματολογίου με μεγαλύτερο λόγο εγκυρότητας περιεχομένου είναι και περισσότερο απαραίτητα να συμπεριληφθούν στο τελικό ερωτηματολόγιο. Το ερώτημα που τίθεται στην περίπτωση αυτή είναι «ποιος είναι ο ελάχιστος λόγος εγκυρότητας περιεχομένου που πρέπει να έχει ένα στοιχείο, έτσι ώστε να συμπεριληφθεί στο τελικό ερωτηματολόγιο;». Στον πίνακα 1 παρουσιάζεται ο ελάχιστος λόγος εγκυρότητας περιεχομένου που πρέπει να έχει ένα στοιχείο, έτσι ώστε να συμπεριληφθεί στο τελικό ερωτηματολόγιο, ανάλογα με τον αριθμό των ειδικών.²⁵ Για παράδειγμα, σύμφωνα με τον πίνακα 1, εάν ο αριθμός των ειδικών που χαρακτηρίζουν τα στοιχεία ενός ερωτηματολογίου είναι 20, τότε τα στοιχεία με ελάχιστο λόγο εγκυρότητας περιεχομένου >0,42 θα συμπεριληφθούν στο τελικό ερωτηματολόγιο, ενώ εάν ο αριθμός των ειδικών είναι 30, τότε τα στοιχεία με ελάχιστο λόγο εγκυρότητας περιεχομένου >0,33 θα συμπεριληφθούν στο τελικό ερωτηματολόγιο.

2.2. Εγκυρότητα εννοιολογικής κατασκευής

Η εγκυρότητα εννοιολογικής κατασκευής (construct validity) αφορά στο βαθμό στον οποίο ένα ερωτηματολόγιο αντανακλά το πραγματικό θεωρητικό νόημα της έννοιας

Πίνακας 1. Ο ελάχιστος λόγος εγκυρότητας περιεχομένου που πρέπει να έχει ένα στοιχείο ενός ερωτηματολογίου, έτσι ώστε να συμπεριληφθεί στο τελικό ερωτηματολόγιο, ανάλογα με τον αριθμό των ειδικών που κρίνουν τα στοιχεία του ερωτηματολογίου.

Αριθμός ειδικών	Ελάχιστος λόγος εγκυρότητας περιεχομένου
5	0,99
6	0,99
7	0,99
8	0,85
9	0,78
10	0,62
11	0,59
12	0,56
13	0,54
14	0,51
15	0,49
20	0,42
25	0,37
30	0,33
35	0,31
40	0,29

την οποία διατείνεται ότι μετρά.^{3,17-23} Η εγκυρότητα εννοιολογικής κατασκευής αφορά κυρίως σε ερωτηματολόγια που χρησιμοποιούνται στις κοινωνικές επιστήμες, στην ψυχολογία και στην εκπαίδευση, για τη μέτρηση ψυχομετρικών εννοιών, πεποιθήσεων και αντιλήψεων. Για παράδειγμα, σε μια μελέτη για την εκτίμηση της καλλιτεχνικής ικανότητας παιδιών του δημοτικού σχολείου, η εγκυρότητα εννοιολογικής κατασκευής αφορά στο βαθμό στον οποίο το ερωτηματολόγιο της μελέτης εκτιμά το πραγματικό θεωρητικό νόημα της καλλιτεχνικής ικανότητας, που αποτελεί μια αφηρημένη και θεωρητική έννοια.

Ο έλεγχος της εγκυρότητας εννοιολογικής κατασκευής είναι έως ένα βαθμό υποκειμενικός και για το λόγο αυτόν απαιτείται η διεξαγωγή σημαντικού αριθμού μελετών σε διαφορετικές χώρες, σε διαφορετικούς μελετώμενους πληθυσμούς και σε διαφορετικές χρονικές στιγμές. Εν τούτοις, σε μια συγκεκριμένη μελέτη, ο έλεγχος της εγκυρότητας εννοιολογικής κατασκευής μπορεί να επιτευχθεί με την παραγοντική ανάλυση, τη συγκλίνουσα εγκυρότητα, τη διακρίνουσα εγκυρότητα και τη μέθοδο των γνωστών ομάδων.

2.2.1. Παραγοντική ανάλυση. Εφαρμόζοντας την παραγοντική ανάλυση (factor analysis) στα στοιχεία ενός ερωτηματολογίου προκύπτουν διάφοροι παράγοντες (factors) που εκφράζουν επί μέρους διαστάσεις της έννοιας που μετρά το ερωτηματολόγιο της μελέτης.^{26,27} Οι παράγοντες αυτοί προκύπτουν με βάση τις συσχετίσεις που παρουσιάζουν μεταξύ τους τα διάφορα στοιχεία ενός ερωτηματολογίου. Για παράδειγμα, εάν ένα ερωτηματολόγιο αποτελείται από 30 στοιχεία, εφαρμόζοντας την παραγοντική ανάλυση προκύπτουν πολύ λιγότεροι παράγοντες, καθένας από τους οποίους περιλαμβάνει ορισμένα από τα 30 στοιχεία του ερωτηματολογίου. Με τον τρόπο αυτόν, μια ευρεία έννοια απλουστεύεται, ομαδοποιείται σε επί μέρους παράγοντες και γίνεται σαφέστερη. Ενδεικτικά αναφέρεται ότι το ερωτηματολόγιο Short Form (36) Health Survey για τη μέτρηση της ποιότητας ζωής αποτελείται από 36 ερωτήσεις που συνθέτουν 8 επί μέρους παράγοντες αξιολόγησης της ποιότητας ζωής (εικ. 1).^{28,29} Καθένας από τους 8 παράγοντες αποτελείται από 2–10 ερωτήσεις. Πιο συγκεκριμένα, οι 8 παράγοντες είναι οι εξής:

- Σωματική λειτουργικότητα (physical functioning), που εκτιμάται με 10 ερωτήσεις
- Σωματικός ρόλος (role physical), που εκτιμάται με 4 ερωτήσεις
- Σωματικός πόνος (bodily pain), που εκτιμάται με 2 ερωτήσεις
- Γενική υγεία (general health), που εκτιμάται με 5 ερωτήσεις



Εικόνα 1. Οι 8 παράγοντες αξιολόγησης της ποιότητας ζωής του Short Form (36) Health Survey.

- Ζωτικότητα (vitality), που εκτιμάται με 4 ερωτήσεις
- Κοινωνική λειτουργικότητα (social functioning), που εκτιμάται με 2 ερωτήσεις
- Συναισθηματικός ρόλος (role emotional), που εκτιμάται με 3 ερωτήσεις
- Ψυχική υγεία (mental health), που εκτιμάται με 5 ερωτήσεις.

Η παραγοντική ανάλυση αποτελεί μια σύνθετη μαθηματική διαδικασία και διακρίνεται στη διερευνητική ανάλυση παραγόντων (explanatory factor analysis) και στην επιβεβαιωτική ανάλυση παραγόντων (confirmatory factor analysis).^{26,27} Η διερευνητική ανάλυση παραγόντων

χρησιμοποιείται στην περίπτωση κατά την οποία κατασκευάζεται ένα νέο ερωτηματολόγιο και οι ερευνητές δεν έχουν εκ των προτέρων γνώση σχετικά με τους παράγοντες που μπορούν να δημιουργηθούν από τα στοιχεία του ερωτηματολογίου. Αντίθετα, η επιβεβαιωτική ανάλυση παραγόντων πραγματοποιείται έπειτα από τη διερευνητική ανάλυση παραγόντων και αφού προηγουμένως έχουν δημιουργηθεί οι επί μέρους παράγοντες, καθώς και τα στοιχεία που τους συνιστούν. Έτσι, στην επιβεβαιωτική ανάλυση παραγόντων, με βάση τα δεδομένα μιας μελέτης, ελέγχεται εάν οι παράγοντες που δημιουργήθηκαν με τη διερευνητική ανάλυση παραγόντων όντως ανταποκρίνονται στην πραγματικότητα. Για παράδειγμα, όπως προαναφέρθηκε, το Short Form (36) Health Survey αποτελείται από 8 επί μέρους παράγοντες. Εάν το Short Form (36) Health Survey χρησιμοποιηθεί σε μια μελέτη στην Ελλάδα, τότε πρέπει να πραγματοποιηθεί επιβεβαιωτική ανάλυση παραγόντων. Στην περίπτωση αυτή, εάν και στη μελέτη στην Ελλάδα προκύψουν οι 8 επί μέρους παράγοντες, με τα αντίστοιχα στοιχεία, που ισχύουν και διεθνώς, τότε το Short Form (36) Health Survey εμφανίζει υψηλή εγκυρότητα εννοιολογικής κατασκευής.

Σημειώνεται ότι στην παραγοντική ανάλυση (α) ο αριθμός των στοιχείων του ερωτηματολογίου πρέπει να είναι 3–5 φορές μεγαλύτερος από τους επί μέρους παράγοντες που θα προκύψουν, (β) ο συνολικός αριθμός των συμμετεχόντων πρέπει να είναι τουλάχιστον >300 και (γ) ο αριθμός των συμμετεχόντων πρέπει να είναι τουλάχιστον 5 φορές μεγαλύτερος (ιδανικά 10 φορές μεγαλύτερος) από τον αριθμό των στοιχείων του ερωτηματολογίου.^{26,27} Για παράδειγμα, εάν οι παράγοντες που προκύψουν σε ένα ερωτηματολόγιο είναι 10, τότε ο αριθμός των στοιχείων του ερωτηματολογίου πρέπει να είναι 30–50. Επί πλέον, εάν ο αριθμός των στοιχείων ενός ερωτηματολογίου είναι 70, τότε πρέπει να υπάρχουν τουλάχιστον 350 συμμετέχοντες.

2.2.2. Συγκλίνουσα εγκυρότητα. Η συγκλίνουσα εγκυρότητα (convergent validity) αφορά στο βαθμό συσχέτισης μεταξύ ενός νέου ερωτηματολογίου και ενός υπάρχοντος, το οποίο έχει εμφανίσει αποδεκτή εγκυρότητα σύμφωνα με τα αποτελέσματα προγενέστερων μελετών.^{3,17–23} Τόσο το νέο ερωτηματολόγιο που κατασκευάζεται σε μια συγκεκριμένη μελέτη όσο και το υπάρχον ερωτηματολόγιο με το οποίο συγκρίνεται μετρούν την ίδια έννοια ή, αλλιώς, την ίδια μεταβλητή. Για παράδειγμα, εάν σε μια μελέτη κατασκευάστηκε ένα νέο ερωτηματολόγιο για την εκτίμηση της ικανότητας των μαθητών δημοτικού στα μαθηματικά, τότε οι βαθμολογίες των μαθητών στο ερωτηματολόγιο αυτό πρέπει να συγκριθούν με τις βαθμολογίες των ίδιων μαθητών σε ένα υπάρχον ερωτηματολόγιο με αποδεδειγμένη εγκυρότητα, που επίσης μετρά την ικανότητα των μαθητών στα μαθηματικά. Σημειώνεται ότι η συμπλήρωση και των

δύο ερωτηματολογίων για τη μέτρηση της ίδιας έννοιας πρέπει να επιτελείται την ίδια χρονική στιγμή ή τουλάχιστον μεταξύ δύο χρονικών στιγμών, μεταξύ των οποίων δεν συμβαίνει κάποια σημαντική αλλαγή, που μπορεί να επηρεάσει τις απαντήσεις των συμμετεχόντων. Στο προαναφερθέν παράδειγμα, τα δύο ερωτηματολόγια για την εκτίμηση της ικανότητας των μαθητών στα μαθηματικά πρέπει να διανεμηθούν την ίδια ημέρα ή έστω σε δύο διαφορετικές ημέρες, χωρίς όμως μεταξύ των δύο αυτών ημερών να έχει συμβεί κάποιο γεγονός (π.χ. μια εκπαιδευτική παρέμβαση) που μπορεί να επηρεάσει τις απαντήσεις των μαθητών. Έτσι, υπολογίζεται ο *συντελεστής συσχέτισης* (correlation coefficient) μεταξύ των βαθμολογιών των μαθητών με τα δύο ερωτηματολόγια, με υψηλές τιμές θετικής συσχέτισης να δηλώνουν την ύπαρξη συγκλίνουσας εγκυρότητας του νέου ερωτηματολογίου. Στο συγκεκριμένο παράδειγμα, θετική συσχέτιση σημαίνει ότι αύξηση της βαθμολογίας στο ένα ερωτηματολόγιο συσχετίζεται με αύξηση της βαθμολογίας και στο άλλο ερωτηματολόγιο.

Το είδος και το μέγεθος της γραμμικής σχέσης μεταξύ δύο ποσοτικών μεταβλητών που ακολουθούν την κανονική κατανομή εκτιμάται με το συντελεστή συσχέτισης του Pearson (Pearson's correlation coefficient), ενώ όταν έστω και μία από τις δύο ποσοτικές μεταβλητές δεν ακολουθεί την κανονική κατανομή, τότε χρησιμοποιείται ο συντελεστής συσχέτισης διατάξεων του Spearman (Spearman's rank correlation coefficient).³⁰ Επί πλέον, ο συντελεστής συσχέτισης διατάξεων του Spearman χρησιμοποιείται και στην περίπτωση κατά την οποία η μία ή και οι δύο μεταβλητές είναι σε διατεταγμένη κλίμακα. Ο συντελεστής συσχέτισης του Kendall (Kendall's tau correlation coefficient), εξ άλλου, χρησιμοποιείται στην περίπτωση που οι δύο μεταβλητές είναι ονομαστικές ή ποσοτικές, οι οποίες δεν ακολουθούν την κανονική κατανομή. Οι συντελεστές συσχέτισης δεν έχουν μονάδα μέτρησης και λαμβάνουν τιμές από -1 έως +1. Οι τιμές -1 και +1 συμβαίνουν όταν υπάρχει μια τέλεια, αρνητική ή θετική αντίστοιχα, γραμμική σχέση μεταξύ των δύο μεταβλητών. Όταν ο συντελεστής συσχέτισης λαμβάνει την τιμή 0 σημαίνει ότι δεν υπάρχει γραμμική σχέση μεταξύ των δύο μεταβλητών, αλλά δεν αποκλείεται η ύπαρξη μιας άλλης μη γραμμικής σχέσης. Εάν το μέγεθος της μιας μεταβλητής τείνει να αυξάνεται όπως αυξάνει το μέγεθος και της άλλης μεταβλητής, τότε υπάρχει *θετική συσχέτιση* (positive correlation) μεταξύ των δύο μεταβλητών και ο συντελεστής συσχέτισης είναι >0. Αντίθετα, εάν το μέγεθος της μιας μεταβλητής τείνει να αυξάνεται όπως ελαττώνεται το μέγεθος της άλλης μεταβλητής, τότε υπάρχει *αρνητική συσχέτιση* (negative correlation) μεταξύ των δύο μεταβλητών και ο συντελεστής συσχέτισης είναι <0. Αναφορικά με τις τιμές που μπορούν να λάβουν οι συντελεστές συσχέτισης, ισχύουν τα εξής:

- Τιμές μεταξύ 0–0,3 ή μεταξύ 0 έως -0,3 υποδηλώνουν μικρή συσχέτιση
- Τιμές μεταξύ 0,31–0,6 ή μεταξύ -0,31 έως -0,6 υποδηλώνουν μέτρια συσχέτιση
- Τιμές μεταξύ 0,61–0,8 ή μεταξύ -0,61 έως -0,8 υποδηλώνουν ισχυρή συσχέτιση
- Τιμές μεταξύ 0,81–1 ή μεταξύ -0,81 έως -1 υποδηλώνουν πολύ ισχυρή συσχέτιση.

2.2.3. Διακρίνουσα εγκυρότητα. Η διακρίνουσα εγκυρότητα (discriminant validity) αφορά στο βαθμό στον οποίο δεν υπάρχει σημαντική συσχέτιση μεταξύ ενός νέου ερωτηματολογίου που μετρά μια συγκεκριμένη έννοια και ενός υπάρχοντος ερωτηματολογίου, το οποίο μετρά μια διαφορετική έννοια και έχει εμφανίσει αποδεκτή εγκυρότητα, σύμφωνα με τα αποτελέσματα προγενέστερων μελετών.^{31–23} Ουσιαστικά, η διακρίνουσα εγκυρότητα είναι η ακριβώς αντίθετη μορφή της συγκλίνουσας εγκυρότητας. Για παράδειγμα, εάν σε μια μελέτη κατασκευάστηκε ένα νέο ερωτηματολόγιο για την εκτίμηση του αυτοελέγχου, τότε οι βαθμολογίες των συμμετεχόντων στο ερωτηματολόγιο αυτό πρέπει να συγκριθούν με τις βαθμολογίες των ίδιων συμμετεχόντων σε ένα υπάρχον ερωτηματολόγιο με αποδεδειγμένη εγκυρότητα, που μετρά όμως μια διαφορετική έννοια, όπως π.χ. την αυτοεκτίμηση. Σημειώνεται ότι η συμπλήρωση των δύο ερωτηματολογίων για τη μέτρηση των δύο διαφορετικών εννοιών πρέπει να πραγματοποιείται την ίδια χρονική στιγμή ή τουλάχιστον μεταξύ δύο χρονικών στιγμών, μεταξύ των οποίων δεν συμβαίνει κάποια σημαντική αλλαγή που μπορεί να επηρεάσει τις απαντήσεις των συμμετεχόντων. Έτσι, υπολογίζεται ο συντελεστής συσχέτισης μεταξύ των βαθμολογιών των συμμετεχόντων με τα δύο διαφορετικά ερωτηματολόγια, με μικρές τιμές συσχέτισης (τιμές που προσεγγίζουν στο μηδέν) να δηλώνουν την ύπαρξη διακρίνουσας εγκυρότητας του νέου ερωτηματολογίου.

2.2.4. Μέθοδος των γνωστών ομάδων. Η μέθοδος των γνωστών ομάδων (known groups method) χρησιμοποιείται για να εκτιμήσει το βαθμό στον οποίο ένα ερωτηματολόγιο μπορεί να διακρίνει δύο διαφορετικές ομάδες ατόμων, που διαφέρουν ως προς κάποιο συγκεκριμένο χαρακτηριστικό.³¹ Για παράδειγμα, εάν σε μια μελέτη χρησιμοποιηθεί ένα ερωτηματολόγιο για την εκτίμηση της κατάθλιψης και διανεμηθεί σε άτομα με κατάθλιψη και σε άτομα χωρίς κατάθλιψη, τότε η βαθμολογία κατάθλιψης στις δύο ομάδες πρέπει να είναι διαφορετική. Θεωρώντας ότι μεγαλύτερες βαθμολογίες δηλώνουν μεγαλύτερο βαθμό κατάθλιψης, θα πρέπει τα άτομα με κατάθλιψη να έχουν μεγαλύτερες βαθμολογίες σε σχέση με τα άτομα χωρίς κατάθλιψη,

έτσι ώστε να θεωρηθεί ότι το ερωτηματολόγιο εμφανίζει εγκυρότητα εννοιολογικής κατασκευής.

2.3. Εγκυρότητα κριτηρίου

Η εγκυρότητα κριτηρίου (criterion validity) αφορά στη χρήση ενός κριτηρίου, ενός υπάρχοντος δηλαδή ερωτηματολογίου με αποδεδειγμένη εγκυρότητα και αξιοπιστία, ως μεθόδου αναφοράς ή χρυσού κανόνα για την εκτίμηση της εγκυρότητας ενός νέου ερωτηματολογίου.^{3,17-23} Τόσο το νέο ερωτηματολόγιο όσο και το υπάρχον πρέπει να μετρούν ακριβώς την ίδια έννοια.

Η εγκυρότητα κριτηρίου χρησιμοποιείται συχνά για τη δημιουργία ενός ερωτηματολογίου μικρότερης έκτασης, με μικρότερο δηλαδή αριθμό στοιχείων, από ένα ερωτηματολόγιο μεγαλύτερης έκτασης με αποδεδειγμένη εγκυρότητα και αξιοπιστία.

Η εγκυρότητα κριτηρίου περιλαμβάνει την ταυτόχρονη εγκυρότητα και την προβλεπτική εγκυρότητα.

2.3.1. Ταυτόχρονη εγκυρότητα. Η ταυτόχρονη εγκυρότητα (concurrent validity) αφορά στο βαθμό συσχέτισης μεταξύ του νέου ερωτηματολογίου και του υπάρχοντος.^{3,17-23} Απαραίτητη προϋπόθεση είναι η συμπλήρωση των δύο ερωτηματολογίων να επιτευχθεί την ίδια χρονική στιγμή. Η ύπαρξη ισχυρής θετικής συσχέτισης μεταξύ του νέου και του υπάρχοντος ερωτηματολογίου δηλώνει υψηλή ταυτόχρονη εγκυρότητα για το νέο ερωτηματολόγιο. Η εκτίμηση της συσχέτισης πραγματοποιείται με τους συντελεστές συσχέτισης που αναλύθηκαν προηγουμένως στην περίπτωση της συγκλίνουσας εγκυρότητας. Για παράδειγμα, το ερωτηματολόγιο Short Form (36) Health Survey για την εκτίμηση της ποιότητας ζωής έχει χρησιμοποιηθεί σε πλήθος μελετών και έχει μεταφραστεί σε δεκάδες γλώσσες. Ο μεγάλος όμως αριθμός στοιχείων (36) που περιλαμβάνει έχει ως αποτέλεσμα σε ορισμένες περιπτώσεις να μειώνεται σημαντικά το ποσοστό απόκρισης των συμμετεχόντων. Σε μια μελέτη, κατασκευάζεται ένα νέο ερωτηματολόγιο με βάση τα στοιχεία του Short Form (36) Health Survey. Το νέο ερωτηματολόγιο περιλαμβάνει μόνο 10 στοιχεία από το ερωτηματολόγιο Short Form (36) Health Survey, οπότε πιστεύεται ότι θα έχει μεγαλύτερο ποσοστό απόκρισης. Στην περίπτωση αυτή, υπολογίζεται ο συντελεστής συσχέτισης μεταξύ της βαθμολογίας ποιότητας ζωής με το νέο σύντομο ερωτηματολόγιο και της βαθμολογίας ποιότητας ζωής με το Short Form (36) Health Survey. Η ύπαρξη ισχυρής θετικής συσχέτισης μεταξύ των δύο βαθμολογιών δηλώνει την ύπαρξη ταυτόχρονης εγκυρότητας του νέου ερωτηματολογίου.

2.3.2. Προβλεπτική εγκυρότητα. Η προβλεπτική εγκυρό-

τητα (predictive validity) αφορά στο βαθμό στον οποίο ένα ερωτηματολόγιο μπορεί να προβλέψει μελλοντικά γεγονότα ή φαινόμενα.^{3,17-23} Στην περίπτωση αυτή, διερευνάται εάν η βαθμολογία σε ένα νέο ερωτηματολόγιο μπορεί να προβλέψει τη βαθμολογία σε ένα υπάρχον ερωτηματολόγιο. Έτσι, οι συμμετέχοντες στη μελέτη πρέπει να συμπληρώσουν το νέο ερωτηματολόγιο σε μια συγκεκριμένη χρονική στιγμή και το υπάρχον ερωτηματολόγιο, που αποτελεί το κριτήριο εκτίμησης της εγκυρότητας, σε μια μετέπειτα χρονική στιγμή. Ακολούθως, υπολογίζεται ο συντελεστής συσχέτισης μεταξύ των δύο ερωτηματολογίων. Η ύπαρξη ισχυρής συσχέτισης μεταξύ των δύο βαθμολογιών δηλώνει την ύπαρξη προβλεπτικής εγκυρότητας του νέου ερωτηματολογίου.

2.4. Εγκυρότητα όψης

Η εγκυρότητα όψης (face validity) δεν αποτελεί μορφή εγκυρότητας με τη στενή έννοια, όπως η εγκυρότητα περιεχομένου, η εγκυρότητα εννοιολογικής κατασκευής και η εγκυρότητα κριτηρίου. Αφορά ουσιαστικά στην εμφάνιση και στη διαμόρφωση ενός ερωτηματολογίου.²⁴ Για παράδειγμα, εάν σε μια μελέτη κατασκευάστηκε ένα νέο ερωτηματολόγιο για την εκτίμηση της ικανότητας των μαθητών δημοτικού στα μαθηματικά και οι μαθητές συμφωνούν ότι το ερωτηματολόγιο είναι κατάλληλο για το σκοπό της μελέτης, τότε το ερωτηματολόγιο θεωρείται ότι είναι αποδεκτό αναφορικά με την εγκυρότητα όψης. Είναι χαρακτηριστικό ότι για τον έλεγχο της εγκυρότητας όψης δεν μπορούν να χρησιμοποιηθούν στατιστικές μέθοδοι, σε αντίθεση με τις υπόλοιπες μορφές εγκυρότητας.

3. ΑΞΙΟΠΙΣΤΙΑ

Με την ευρεία έννοια, όταν σε μια μέτρηση δεν υπάρχει τυχαίο σφάλμα (random error), τότε η μέτρηση λέγεται ότι έχει αξιοπιστία, ακρίβεια (precision), αναπαραγωγιμότητα (reproducibility), επαναληπτικότητα (repeatability) ή συνέπεια (consistency).^{8,9} Η παρουσία τυχαίου σφάλματος καθιστά τη μέτρηση αναξιόπιστη, ανακριβή κ.ο.κ. Τυχαίο σφάλμα μέτρησης καλείται η διαφορά μεταξύ της εμπειρικής (παρατηρήσιμης) τιμής μιας μεταβλητής που προκύπτει έπειτα από μια μέτρηση και της μέσης τιμής των εμπειρικών τιμών που προκύπτουν έπειτα από ένα σύνολο μετρήσεων.^{8,9} Η επανάληψη της μέτρησης είναι απαραίτητη για τη διαπίστωση των τυχαίων σφαλμάτων. Για παράδειγμα, εάν το βάρος ενός ατόμου μετρηθεί με μια ζυγαριά δέκα φορές και προκύψει ίσο με 80 kg και στις δέκα μετρήσεις, τότε η ζυγαριά αυτή είναι αξιόπιστη ή, αλλιώς, ακριβής. Όσο μεγαλύτερη είναι η μεταβλητότητα των δέκα μετρήσεων του βάρους, τόσο

μικρότερη είναι η αξιοπιστία της ζυγαριάς. Εάν, εξάλλου, θεωρηθεί μια άλλη ζυγαριά ως μέθοδος αναφοράς ή χρυσός κανόνας, σύμφωνα με την οποία το πραγματικό βάρος του ατόμου είναι ίσο με 80 kg, τότε η πρώτη ζυγαριά με την οποία πραγματοποιήθηκαν οι δέκα μετρήσεις του βάρους θεωρείται έγκυρη. Όσο μεγαλύτερη είναι η απόσταση μεταξύ του μέσου βάρους που προέκυψε από τις δέκα μετρήσεις της πρώτης ζυγαριάς και του πραγματικού βάρους που προέκυψε από τη δεύτερη ζυγαριά που θεωρείται ως μέθοδος αναφοράς, τόσο μικρότερη είναι η εγκυρότητα της ζυγαριάς. Είναι σαφές ότι μια μέθοδος μέτρησης μπορεί να εμφανίζει τέλεια αξιοπιστία, να εμφανίζει δηλαδή σε όλες τις μετρήσεις την ίδια τιμή για μια μεταβλητή, χωρίς όμως αυτό να συνεπάγεται απαραίτητα και τέλεια εγκυρότητα, καθώς όλες οι μετρήσεις της μεταβλητής είναι δυνατόν να απέχουν από την πραγματική της τιμή.

Η αξιοπιστία ή, αλλιώς, η ακρίβεια ενός ερωτηματολογίου αφορά στη σταθερότητα ή, αλλιώς, στη συνέπεια με την οποία το ερωτηματολόγιο μετρά την έννοια ή, αλλιώς, τη μεταβλητή την οποία διατείνεται ότι μετρά.^{7,10,15,17-23} Η αύξηση της αξιοπιστίας ενός ερωτηματολογίου συνεπάγεται τη μείωση του τυχαίου σφάλματος. Για την εκτίμηση του τυχαίου σφάλματος δεν απαιτείται η γνώση της πραγματικής τιμής της έννοιας που μετράται. Σημειώνεται ότι η αξιοπιστία ενός ερωτηματολογίου αφορά στις απαντήσεις των συμμετεχόντων και όχι στο ερωτηματολόγιο καθ' αυτό. Για το λόγο αυτόν, η αξιοπιστία ενός ερωτηματολογίου πρέπει να ελέγχεται πάντοτε σε μια μελέτη, ανεξάρτητα από το αν το ερωτηματολόγιο έχει εμφανίσει υψηλή αξιοπιστία σε προγενέστερες μελέτες.

Η εκτίμηση της αξιοπιστίας ενός ερωτηματολογίου περιλαμβάνει την εκτίμηση της αξιοπιστίας ελέγχου-επανελέγχου, της αξιοπιστίας εναλλακτικών μορφών, της αξιοπιστίας ημίσεων τμημάτων, της αξιοπιστίας εσωτερικής συνέπειας και της αξιοπιστίας μεταξύ παρατηρητών.

3.1. Αξιοπιστία ελέγχου-επανελέγχου

Η αξιοπιστία ελέγχου-επανελέγχου (test-retest reliability) εκτιμά το βαθμό συσχέτισης των απαντήσεων των συμμετεχόντων σε ένα ερωτηματολόγιο σε δύο διαφορετικές χρονικές στιγμές.^{7,10,15,17-23} Πιο συγκεκριμένα, για την εκτίμηση της αξιοπιστίας ελέγχου-επανελέγχου πραγματοποιείται η εξής διαδικασία:

- Το ερωτηματολόγιο συμπληρώνεται από 30–50 συμμετέχοντες στο πλαίσιο μιας πιλοτικής μελέτης
- Έπειτα από την πάροδο ενός ορισμένου χρονικού διαστήματος, το ερωτηματολόγιο συμπληρώνεται από τους ίδιους συμμετέχοντες

- Υπολογίζεται ο συντελεστής συσχέτισης των απαντήσεων των συμμετεχόντων στις δύο διαφορετικές χρονικές στιγμές

- Η ύπαρξη ισχυρής θετικής συσχέτισης δηλώνει την ύπαρξη αξιοπιστίας ελέγχου-επανελέγχου.

Απαιτείται ιδιαίτερη προσοχή, έτσι ώστε η συμπλήρωση του ερωτηματολογίου στις διαφορετικές χρονικές στιγμές να πραγματοποιείται κάτω από ακριβώς τις ίδιες συνθήκες. Για παράδειγμα, η εκτίμηση της αξιοπιστίας ελέγχου-επανελέγχου ενός ερωτηματολογίου αναφορικά με τις γνώσεις των επαγγελματιών υγείας για τη διαχείριση των μαζικών καταστροφών δεν μπορεί να επιτευχθεί πριν και μετά την πραγματοποίηση ενός αντίστοιχου εκπαιδευτικού προγράμματος, καθώς οι γνώσεις των επαγγελματιών υγείας αναμένεται να αυξηθούν έπειτα από την εφαρμογή του προγράμματος. Είναι σαφές ότι για την εκτίμηση της αξιοπιστίας ελέγχου-επανελέγχου δεν πρέπει να έχει συμβεί κάποια σημαντική μεταβολή μεταξύ των διαφορετικών χρονικών στιγμών, που μπορεί να επηρεάσει τις απαντήσεις των συμμετεχόντων.

Επί πλέον, ιδιαίτερης σημασίας είναι το χρονικό διάστημα που μεσολαβεί μεταξύ της συμπλήρωσης του ερωτηματολογίου από τους συμμετέχοντες. Συνήθως, όσο μικρότερο είναι το χρονικό διάστημα που μεσολαβεί μεταξύ της συμπλήρωσης του ερωτηματολογίου, τόσο μεγαλύτερη είναι η συσχέτιση μεταξύ των απαντήσεων των συμμετεχόντων. Όταν το χρονικό διάστημα που μεσολαβεί μεταξύ της συμπλήρωσης του ερωτηματολογίου είναι σχετικά μικρό, τότε υπάρχει αυξημένη πιθανότητα να εξοικειωθούν οι συμμετέχοντες με το ερωτηματολόγιο και, βασιζόμενοι στη μνήμη τους, να δώσουν τις ίδιες απαντήσεις. Πρόκειται για μια κατάσταση που είναι γνωστή ως «αποτέλεσμα μνήμης» (memory effect) και την οποία πρέπει να λαμβάνουν σοβαρά υπ' όψη τους οι ερευνητές στον καθορισμό του χρονικού διαστήματος που μεσολαβεί μεταξύ της συμπλήρωσης του ερωτηματολογίου.

3.2. Αξιοπιστία εναλλακτικών μορφών

Η αξιοπιστία εναλλακτικών μορφών (alternative forms reliability) ή, αλλιώς, αξιοπιστία παρόμοιων μορφών (parallel-forms reliability) χρησιμοποιείται για να εξουδετερώσει την επίδραση του «αποτελέσματος μνήμης» που υπεισέρχεται στην εκτίμηση της αξιοπιστίας ελέγχου-επανελέγχου.^{7,10,15,17-23}

Για την εκτίμηση της αξιοπιστίας εναλλακτικών μορφών, αρχικά, πρέπει να δημιουργηθούν δύο εναλλακτικές ή, αλλιώς, παρόμοιες μορφές του ίδιου ερωτηματολογίου. Ένας τρόπος για να επιτευχθεί αυτό είναι να δημιουργηθεί ένα ερωτηματολόγιο με μεγάλο αριθμό στοιχείων που να

μετρούν την ίδια έννοια. Στη συνέχεια, τα στοιχεία του ερωτηματολογίου κατανέμονται με τυχαίο τρόπο σε δύο ομάδες, που αποτελούν τις δύο εναλλακτικές μορφές του ίδιου ερωτηματολογίου. Οι δύο αυτές εναλλακτικές μορφές συμπληρώνονται από τους ίδιους συμμετέχοντες και στη συνέχεια υπολογίζεται ο συντελεστής συσχέτισης. Η ύπαρξη ισχυρής θετικής συσχέτισης δηλώνει την ύπαρξη αξιοπιστίας εναλλακτικών μορφών. Το πρόβλημα στην περίπτωση αυτή συνίσταται στο γεγονός ότι είναι εξαιρετικά δύσκολο να δημιουργηθεί ένας μεγάλος αριθμός στοιχείων που να μετρούν με πανομοιότυπο τρόπο την ίδια ακριβώς έννοια. Επί πλέον, στην προσέγγιση αυτή, απαραίτητη προϋπόθεση για την ασφαλή εκτίμηση της αξιοπιστίας εναλλακτικών μορφών αποτελεί οι δύο εναλλακτικές μορφές που δημιουργούνται να είναι παρόμοιες ή, αλλιώς, ισοδύναμες, γεγονός που είναι εξαιρετικά δύσκολο στην πράξη. Για το λόγο αυτόν, η εκτίμηση της αξιοπιστίας εναλλακτικών μορφών, συνήθως, πραγματοποιείται με την εξής διαδικασία:

- Το ερωτηματολόγιο με μια συγκεκριμένη μορφή στοιχείων και απαντήσεων συμπληρώνεται από 30–50 συμμετέχοντες στο πλαίσιο μιας πιλοτικής μελέτης
- Δημιουργείται μια εναλλακτική μορφή του ερωτηματολογίου, χωρίς όμως να μεταβάλλεται το νοηματικό περιεχόμενο των στοιχείων αυτού. Συνήθως, μεταβάλλεται σε μικρό βαθμό η διατύπωση των στοιχείων του ερωτηματολογίου ή μεταβάλλεται η σειρά των απαντήσεων στα διάφορα στοιχεία. Έτσι, προκύπτει μια εναλλακτική μορφή του ερωτηματολογίου, παρόμοια βεβαίως με την αρχική μορφή του ερωτηματολογίου αναφορικά με το νοηματικό περιεχόμενο
- Έπειτα από την πάροδο ενός ορισμένου χρονικού διαστήματος, η εναλλακτική μορφή του ερωτηματολογίου συμπληρώνεται από τους ίδιους συμμετέχοντες
- Υπολογίζεται ο συντελεστής συσχέτισης των απαντήσεων των συμμετεχόντων μεταξύ της αρχικής και της εναλλακτικής μορφής του ερωτηματολογίου
- Η ύπαρξη ισχυρής θετικής συσχέτισης δηλώνει την ύπαρξη αξιοπιστίας ελέγχου-επανελέγχου.

Για παράδειγμα, σε ένα ερωτηματολόγιο για την ικανοποίηση των επαγγελματιών υγείας από την εργασία τους σε δημόσια νοσοκομεία, για την εκτίμηση της αξιοπιστίας εναλλακτικών μορφών, στην εναλλακτική μορφή του ερωτηματολογίου η διατύπωση των στοιχείων παρέμεινε η ίδια, αλλά μεταβλήθηκε η σειρά των απαντήσεων στα διάφορα στοιχεία. Ενδεικτικά, στον πίνακα 2 παρατίθεται η μεταβολή της σειράς των απαντήσεων αναφορικά με ένα στοιχείο του ερωτηματολογίου για την εκτίμηση της

Πίνακας 2. Αρχική και εναλλακτική μορφή ενός στοιχείου ενός ερωτηματολογίου για την ικανοποίηση των επαγγελματιών υγείας από την εργασία τους σε δημόσια νοσοκομεία, για την εκτίμηση της αξιοπιστίας εναλλακτικών μορφών. Στην περίπτωση αυτή, μεταβλήθηκε η σειρά των απαντήσεων.

Αρχική μορφή του ερωτηματολογίου				
Πόσο ικανοποιημένος(η) είστε από το ωράριο εργασίας σας;				
Καθόλου	Λίγο	Μέτρια	Αρκετά	Πολύ
1	2	3	4	5
Εναλλακτική μορφή του ερωτηματολογίου				
Πόσο ικανοποιημένος(η) είστε από το ωράριο εργασίας σας;				
Πολύ	Αρκετά	Μέτρια	Λίγο	Καθόλου
5	4	3	2	1

ικανοποίησης των επαγγελματιών υγείας από την εργασία τους σε δημόσια νοσοκομεία. Στον πίνακα 2, η ύπαρξη αξιοπιστίας εναλλακτικών μορφών δηλώνεται από τις παρόμοιες απαντήσεις των συμμετεχόντων στην αρχική και στην εναλλακτική μορφή του ερωτηματολογίου. Οι συμμετέχοντες, δηλαδή, που απάντησαν «πολύ» στην αρχική μορφή του ερωτηματολογίου, θα πρέπει να απαντήσουν «πολύ» και στην εναλλακτική μορφή του ερωτηματολογίου, οι συμμετέχοντες που απάντησαν «αρκετά» στην αρχική μορφή του ερωτηματολογίου, θα πρέπει να απαντήσουν «αρκετά» και στην εναλλακτική μορφή του ερωτηματολογίου κ.ο.κ.

3.3. Αξιοπιστία ημίσεων τμημάτων

Η αξιοπιστία ημίσεων τμημάτων (split-half reliability) εκτιμά το βαθμό συσχέτισης των απαντήσεων των συμμετεχόντων μεταξύ δύο επί μέρους τμημάτων του ίδιου ερωτηματολογίου.^{7,10,15,17–23} Πιο συγκεκριμένα, για την εκτίμηση της αξιοπιστίας ημίσεων τμημάτων πραγματοποιείται η εξής διαδικασία:

- Το ερωτηματολόγιο συμπληρώνεται από 30–50 συμμετέχοντες στο πλαίσιο μιας πιλοτικής μελέτης
- Τα στοιχεία του ερωτηματολογίου διαιρούνται σε δύο επί μέρους τμήματα. Συνήθως, η διαίρεση των στοιχείων πραγματοποιείται με βάση το αν αντιστοιχούν σε άρτιο ή περιττό αριθμό. Για παράδειγμα, εάν ένα ερωτηματολόγιο αποτελείται από δέκα στοιχεία, τότε τα στοιχεία που αντιστοιχούν σε άρτιους αριθμούς, δηλαδή τα στοιχεία που αντιστοιχούν στους αριθμούς 2, 4, 6, 8 και 10, αποτελούν το ένα επί μέρους τμήμα του ερωτηματολογίου, ενώ τα στοιχεία που αντιστοιχούν σε περιττούς αριθμούς, δηλαδή τα στοιχεία που αντιστοιχούν στους αριθμούς 1, 3, 5, 7 και 9, αποτελούν

το δεύτερο επί μέρους τμήμα του ερωτηματολογίου

- Υπολογίζεται η βαθμολογία των συμμετεχόντων στα δύο επί μέρους τμήματα του ερωτηματολογίου
- Υπολογίζεται ο συντελεστής συσχέτισης των βαθμολογιών των συμμετεχόντων μεταξύ των δύο επί μέρους τμημάτων του ερωτηματολογίου
- Η ύπαρξη ισχυρής θετικής συσχέτισης δηλώνει την ύπαρξη αξιοπιστίας ημίσεων τμημάτων.

Για παράδειγμα, για την εκτίμηση της αξιοπιστίας ημίσεων τμημάτων ενός ερωτηματολογίου με 20 στοιχεία αναφορικά με την ποιότητα ζωής πασχόντων με σακχαρώδη διαβήτη, το ερωτηματολόγιο διαιρείται σε δύο επί μέρους τμήματα με δέκα στοιχεία το καθένα, με βάση το αν τα στοιχεία αντιστοιχούν σε άρτιο ή περιττό αριθμό. Έπειτα, υπολογίζεται η βαθμολογία της ποιότητας ζωής στα δύο επί μέρους τμήματα του ερωτηματολογίου. Ακολούθως, υπολογίζεται ο συντελεστής συσχέτισης των βαθμολογιών των συμμετεχόντων μεταξύ των δύο επί μέρους τμημάτων του ερωτηματολογίου. Τιμές του συντελεστή συσχέτισης που κυμαίνονται από 0,61–0,8 υποδηλώνουν ισχυρή θετική συσχέτιση, ενώ τιμές μεταξύ 0,81–1 υποδηλώνουν πολύ ισχυρή θετική συσχέτιση και κατ' επέκταση αποδεκτή αξιοπιστία ημίσεων τμημάτων

3.4. Αξιοπιστία εσωτερικής συνέπειας

Η αξιοπιστία εσωτερικής συνέπειας (internal consistency reliability) εκτιμά τη συνέπεια των απαντήσεων των συμμετεχόντων στα στοιχεία ενός ερωτηματολογίου, εκτιμά δηλαδή εάν οι συμμετέχοντες απαντούν με παρόμοιο τρόπο στα στοιχεία του ερωτηματολογίου.^{7,10,15,17–23} Για παράδειγμα, εάν ένας συμμετέχοντας συμφωνήσει με τη δήλωση «μου αρέσει να παίζω ποδόσφαιρο» και διαφωνήσει με τη δήλωση «απεχθάνομαι το ποδόσφαιρο», τότε υπάρχει συνέπεια στις απαντήσεις του και το ερωτηματολόγιο εμφανίζει αποδεκτή αξιοπιστία εσωτερικής συνέπειας.

Εάν οι απαντήσεις στα στοιχεία ενός ερωτηματολογίου αποτελούνται από δύο κατηγορίες (π.χ. «ναι» και «όχι» ή «σωστό» και «λάθος»), τότε η αξιοπιστία εσωτερικής συνέπειας εκτιμάται με το συντελεστή Kuder-Richardson, ενώ εάν οι απαντήσεις στα στοιχεία ενός ερωτηματολογίου αποτελούνται από >2 κατηγορίες (όπως π.χ., στην περίπτωση των κλιμάκων Likert), τότε η αξιοπιστία εσωτερικής συνέπειας εκτιμάται με το συντελεστή Cronbach's alpha.^{32,33} Οι τιμές του συντελεστή Kuder-Richardson και του συντελεστή Cronbach's alpha πρέπει να είναι τουλάχιστον $\geq 0,7$ (πίν. 3). Σύμφωνα με τα δεδομένα του πίνακα 3, η αξιοπιστία εσωτερικής συνέπειας είναι άριστη όταν

Πίνακας 3. Ερμηνεία των τιμών των συντελεστών Kuder-Richardson και Cronbach's alpha, που χρησιμοποιούνται για την εκτίμηση της αξιοπιστίας εσωτερικής συνέπειας.

Τιμή του συντελεστή	Αξιοπιστία εσωτερικής συνέπειας
<0,5	Μη αποδεκτή
0,5–0,59	Πτωχή
0,6–0,69	Αμφισβητήσιμη
0,7–0,79	Αποδεκτή
0,8–0,89	Καλή
0,9–0,94	Άριστη

ο συντελεστής Kuder-Richardson ή ο συντελεστής Cronbach's alpha λαμβάνουν τιμές 0,9–0,94. Σημειώνεται ότι τιμές $\geq 0,95$ αντιμετωπίζονται με δυσπιστία και δεν είναι επιθυμητές, καθώς δηλώνουν ότι τα διάφορα στοιχεία ενός ερωτηματολογίου είναι ακριβώς όμοια μεταξύ τους και ουσιαστικά αποτελούν επανάληψη, μη προσφέροντας ξεχωριστή πληροφορία το καθένα, αλλά ακριβώς την ίδια πληροφορία. Όταν ο συντελεστής Kuder-Richardson ή ο συντελεστής Cronbach's alpha λαμβάνουν μη αποδεκτές τιμές, δηλαδή τιμές $< 0,7$ ή $\geq 0,95$, τότε καθίσταται αναγκαία η τροποποίηση ή η αφαίρεση μερικών στοιχείων, έτσι ώστε να προκύψουν αποδεκτές τιμές. Σημειώνεται ότι ο αριθμός των στοιχείων που μπορούν να αφαιρεθούν δεν πρέπει να υπερβαίνει το 20% των στοιχείων του αρχικού ερωτηματολογίου. Για παράδειγμα, εάν το αρχικό ερωτηματολόγιο αποτελείται από δέκα στοιχεία, τότε μπορούν να αφαιρεθούν μέχρι και δύο στοιχεία.

3.5. Αξιοπιστία μεταξύ παρατηρητών

Στην περίπτωση κατά την οποία δύο ή περισσότεροι παρατηρητές ή βαθμολογητές αξιολογούν ή βαθμολογούν μια μεταβλητή, το ερώτημα που τίθεται είναι εάν τα αποτελέσματα της συγκεκριμένης διαδικασίας εμφανίζουν αξιοπιστία ή, αλλιώς, συνέπεια. Η αξιοπιστία μεταξύ παρατηρητών (inter-observers reliability) ή, αλλιώς, η αξιοπιστία μεταξύ βαθμολογητών (inter-raters reliability) εκτιμά το βαθμό συμφωνίας μεταξύ των απαντήσεων των παρατηρητών ή των βαθμολογητών. Για την εκτίμηση της αξιοπιστίας μεταξύ παρατηρητών/βαθμολογητών χρησιμοποιούνται ο δείκτης κάρππα και ο συντελεστής συσχέτισης εντός ομάδων.

3.5.1. Δείκτης κάρππα. Εάν οι απαντήσεις στα στοιχεία ενός ερωτηματολογίου αποτελούνται από κατηγορίες, τότε η εκτίμηση της αξιοπιστίας μεταξύ παρατηρητών/βαθμολογητών πραγματοποιείται με αξιολογήσεις από δύο τουλάχιστον διαφορετικούς παρατηρητές/βαθμολογητές

και την εκτίμηση του ποσοστού συμφωνίας μεταξύ τους. Στην περίπτωση αυτή, υπολογίζεται ο δείκτης κάππα (kappa index) ή, αλλιώς, ο συντελεστής κάππα (kappa coefficient) σύμφωνα με την εξής ισότητα:⁸

$$\text{Δείκτης κάππα} = \frac{O - C}{1 - C} \quad (2)$$

Στην ισότητα 2, με O (observed) συμβολίζεται το ποσοστό της παρατηρηθείσας συμφωνίας, με C (chance) συμβολίζεται το ποσοστό της τυχαίας συμφωνίας και 1 είναι το ποσοστό της πλήρους συμφωνίας μεταξύ των παρατηρητών/βαθμολογητών.

Ο υπολογισμός του δείκτη κάππα δεν προϋποθέτει τη γνώση της πραγματικής τιμής του μεγέθους που μετράται. Συνδέει το επίπεδο της παρατηρηθείσας συμφωνίας με το επίπεδο της τυχαίας συμφωνίας και εκτιμά τη μεταβλητότητα που υπάρχει σε κάθε παρατηρητή/βαθμολογητή (within observer/rater variation) και η οποία προκύπτει όταν ο ίδιος παρατηρητής/βαθμολογητής αξιολογεί διαφορετικά σε επανειλημμένες αξιολογήσεις το ίδιο μέγεθος.

Η μέγιστη τιμή του δείκτη κάππα είναι 1 και αντιπροσωπεύει την πλήρη συμφωνία και κατ'επέκταση τη μέγιστη αξιοπιστία μεταξύ παρατηρητών/βαθμολογητών, ενώ ο δείκτης κάππα λαμβάνει την τιμή 0 όταν υπάρχει μόνο τυχαία συμφωνία μεταξύ των παρατηρητών/βαθμολογητών και κατ'επέκταση μηδενική αξιοπιστία μεταξύ παρατηρητών/βαθμολογητών. Οι αρνητικές τιμές είναι μαθηματικά δυνατόν να προκύψουν αντιπροσωπεύοντας συμφωνία μικρότερη από την τυχαία, αλλά είναι σχεδόν απίθανο να εμφανιστούν στη συνήθη πρακτική. Στον πίνακα 4 φαίνεται η ερμηνεία των τιμών του δείκτη κάππα στις μελέτες που διερευνούν την αξιοπιστία μεταξύ παρατηρητών/βαθμολογητών.

Στον πίνακα 5 φαίνονται τα αποτελέσματα της αξιολόγησης 60 ακτινογραφιών από δύο ακτινολόγους με περίπου το ίδιο επίπεδο γνώσεων. Η απόφαση των ακτινολόγων ότι η ακτινογραφία είναι «θετική» ή «αρνητική» αναφορικά με μια συγκεκριμένη πάθηση δεν είναι ούτε αληθής ούτε ψευδής,

Πίνακας 4. Ερμηνεία των τιμών του δείκτη κάππα στις μελέτες που διερευνούν την αξιοπιστία μεταξύ παρατηρητών/βαθμολογητών.

Τιμή του δείκτη κάππα	Βαθμός αξιοπιστίας
0,00	Μηδενική
0,01–0,20	Ελάχιστη
0,21–0,40	Μικρή
0,41–0,60	Μέτρια
0,61–0,80	Σημαντική
≥0,80	Μέγιστη

Πίνακας 5. Τα αποτελέσματα της αξιολόγησης (θετικής ή αρνητικής) 60 ακτινογραφιών από δύο ακτινολόγους με περίπου το ίδιο επίπεδο γνώσεων.

		1ος εξεταστής		Σύνολο
		Θετική αξιολόγηση (+)	Αρνητική αξιολόγηση (-)	
2ος εξεταστής	Θετική αξιολόγηση (+)	11	3	14 (n ₁)
	Αρνητική αξιολόγηση (-)	11	35	46 (n ₂)
	Σύνολο	22 (n ₃)	38 (n ₄)	60 (n)

καθώς η πραγματική τιμή (παρουσία ή μη της πάθησης) δεν είναι γνωστή. Οι δύο ακτινολόγοι συμφωνούν ότι οι 11 από τις 60 ακτινογραφίες είναι θετικές (υποδηλώνοντας την παρουσία της μελετώμενης πάθησης) και ότι οι 35 από τις 60 ακτινογραφίες είναι αρνητικές (υποδηλώνοντας την απουσία της πάθησης). Το ποσοστό της παρατηρηθείσας συμφωνίας μεταξύ των δύο ακτινολόγων, με βάση τα δεδομένα του πίνακα 5, είναι:

$$O = \frac{11 + 35}{60} = 0,77$$

Στον πίνακα 6 φαίνονται τα αποτελέσματα αξιολόγησης των 60 ακτινογραφιών από τους δύο ακτινολόγους που οφείλονται αποκλειστικά και μόνο στην τύχη, οπότε το ποσοστό της τυχαίας συμφωνίας είναι:

$$C = \frac{5,13 + 29,13}{60} = 0,57$$

Έτσι, με βάση την ισότητα 2, ο δείκτης κάππα ισούται με:

$$\text{Δείκτης κάππα} = \frac{O - C}{1 - C} = \frac{0,77 + 0,57}{1 - 0,57} = 0,47$$

Η τιμή του δείκτη κάππα είναι ίση με 0,47, γεγονός που, με βάση τον πίνακα 4, δηλώνει μέτρια αξιοπιστία μεταξύ των δύο ακτινολόγων.

Σημειώνεται ότι εάν οι απαντήσεις στα στοιχεία ενός ερωτηματολογίου αποτελούνται από κατηγορίες, που είναι όμως διατεταγμένες με κάποια σειρά, όπως π.χ. στην περίπτωση των παθολογοανατόμων που χαρακτηρίζουν ένα κυτταρολογικό δείγμα ως φυσιολογικό, ως δείγμα με ελαφρά ατυπία, ως δείγμα με σοβαρή ατυπία και ως καρκινικό, τότε, εκτός από την πλήρη συμφωνία, είναι χρήσιμο να συνυπολογιστεί και η μερική συμφωνία. Στην περίπτωση αυτή, λαμβάνονται υπ' όψη οι διαφορές ανάμεσα στις επί μέρους αξιολογήσεις ανάλογα με το πόσο απέχουν αυτές μεταξύ τους και υπολογίζεται ο σταθμισμένος δείκτης κάππα (weighted kappa coefficient) που συνεκτιμά την απόσταση μεταξύ των αξιολογήσεων. Για παράδειγμα, η

Πίνακας 6. Τα αποτελέσματα της αξιολόγησης (θετικής ή αρνητικής) 60 ακτινογραφιών από τους δύο ακτινολόγους που οφείλονται αποκλειστικά και μόνο στην τύχη.

	1ος εξεταστής		Σύνολο
	Θετική αξιολόγηση (+)	Αρνητική αξιολόγηση (-)	
2ος εξεταστής			
Θετική αξιολόγηση (+)	5,13 $\left(= \frac{n_1 \times n_3}{n} \right)$	8,87 $\left(= \frac{n_1 \times n_4}{n} \right)$	14 (n ₁)
Αρνητική αξιολόγηση (-)	16,87 $\left(= \frac{n_2 \times n_3}{n} \right)$	29,13 $\left(= \frac{n_2 \times n_4}{n} \right)$	46 (n ₂)
Σύνολο	22 (n ₃)	38 (n ₄)	60 (n)

απόσταση μεταξύ ελαφράς και σοβαρής ατυπίας είναι μικρότερη από την απόσταση μεταξύ φυσιολογικού δείγματος και καρκινικού.

3.5.2. *Συντελεστής συσχέτισης εντός ομάδων.* Εάν οι απαντήσεις στα στοιχεία ενός ερωτηματολογίου είναι με τη μορφή ποσοτικών μεταβλητών (όπως π.χ. οι διάφορες βαθμολογίες), τότε η εκτίμηση της αξιοπιστίας μεταξύ παρατηρητών/βαθμολογητών πραγματοποιείται με το συντελεστή συσχέτισης εντός ομάδων (intraclass correlation coefficient).

Ο συντελεστής συσχέτισης εντός ομάδων έχει τα ίδια χαρακτηριστικά με τους υπόλοιπους συντελεστές συσχέτισης και τα οποία αναλύθηκαν στην ενότητα της συγκλίνουσας εγκυρότητας. Υψηλές τιμές του συντελεστή συσχέτισης εντός ομάδων υποδηλώνουν την ύπαρξη αξιοπιστίας μεταξύ παρατηρητών/βαθμολογητών. Ως παράδειγμα εφαρμογής του συντελεστή συσχέτισης εντός ομάδων αναφέρεται μια μελέτη στην οποία παιδιά με σακχαρώδη διαβήτη και οι γονείς τους συμπλήρωσαν ένα ερωτηματολόγιο για την εκτίμηση της ποιότητας ζωής. Έτσι, προέκυψαν οι βαθμολογίες αναφορικά με την ποιότητα ζωής τόσο των παιδιών όσο και των γονιών τους. Έπειτα, υπολογίστηκε ο συντελεστής συσχέτισης εντός ομάδων των βαθμολογιών μεταξύ των παιδιών και των γονιών τους.

4. ΕΠΙ ΜΕΡΟΥΣ ΕΝΟΤΗΤΕΣ ΕΝΟΣ ΕΡΩΤΗΜΑΤΟΛΟΓΙΟΥ

Επισημαίνεται ότι ένα ερωτηματολόγιο σε μια επιδημιολογική μελέτη συνήθως αποτελείται από περισσότερες από μία ενότητες, κάθε μια από τις οποίες μετρά διαφορετικές διαστάσεις ή έννοιες. Για παράδειγμα, σε μια συγχρονική μελέτη εκτίμησης του επιπολασμού διαφόρων δημογραφικών χαρακτηριστικών των κατοίκων μιας πόλης, το ερωτηματολόγιο της μελέτης θα περιλαμβάνει ερωτήσεις αναφορικά μόνο με τα δημογραφικά χαρακτηριστικά. Όμως, σε μια συγχρονική μελέτη για τη διερεύνηση της σχέσης μεταξύ

δημογραφικών χαρακτηριστικών και ποιότητας ζωής, θα χρησιμοποιηθεί ένα ερωτηματολόγιο με δύο διαφορετικές ενότητες: (α) η πρώτη ενότητα θα περιλαμβάνει ερωτήσεις αναφορικά με τα δημογραφικά χαρακτηριστικά και (β) η δεύτερη ενότητα θα περιέχει ερωτήσεις για την εκτίμηση της ποιότητας ζωής.

Είναι σαφές ότι οι διάφορες ενότητες ενός ερωτηματολογίου που είναι τελείως διαφορετικές μεταξύ τους αποτελούν ουσιαστικά επί μέρους ξεχωριστά ερωτηματολόγια, για καθένα από τα οποία πρέπει να γίνει ξεχωριστά έλεγχος της εγκυρότητας και της αξιοπιστίας. Για παράδειγμα, εάν το ερωτηματολόγιο που χρησιμοποιείται σε μια μελέτη αποτελείται από τρεις διαφορετικές ενότητες (η πρώτη ενότητα αφορά στα δημογραφικά χαρακτηριστικά, η δεύτερη ενότητα αφορά στην ποιότητα ζωής και η τρίτη ενότητα αφορά στην ικανοποίηση από τις παρεχόμενες υπηρεσίες υγείας), τότε ουσιαστικά χρησιμοποιούνται τρία επί μέρους ξεχωριστά ερωτηματολόγια, για καθένα από τα οποία πρέπει να πραγματοποιηθεί έλεγχος της εγκυρότητας και της αξιοπιστίας. Προηγουμένως αναλύθηκε εκτενώς ο τρόπος με τον οποίο ελέγχεται η εγκυρότητα και η αξιοπιστία των διαφόρων ενότητων ενός ερωτηματολογίου. Σημειώνεται ότι για κάθε ενότητα του ερωτηματολογίου, ο έλεγχος της εγκυρότητας και της αξιοπιστίας δεν είναι απαραίτητο να περιλαμβάνει όλα όσα προαναφέρθηκαν στις ενότητες της εγκυρότητας και της αξιοπιστίας. Κάθε ενότητα μπορεί να έχει ερωτήσεις ή δηλώσεις/προτάσεις διαφορετικής μορφής και ανάλογα με την περίπτωση εφαρμόζεται ο κατάλληλος έλεγχος εγκυρότητας και αξιοπιστίας. Για παράδειγμα, στην ενότητα ενός ερωτηματολογίου που περιλαμβάνει τα δημογραφικά χαρακτηριστικά, είναι δυνατόν να εκτιμηθεί μόνο η εγκυρότητα όψης και η αξιοπιστία ελέγχου-επανελέγχου.

Χρειάζεται ιδιαίτερη προσοχή στην περίπτωση κατά την οποία ένα ερωτηματολόγιο αποτελείται από περισσότερες από μία ενότητες. Στην περίπτωση αυτή, οι ερωτήσεις κάθε ενότητας θα πρέπει να είναι σαφώς καθορισμένες και ευδιάκριτες από τις ερωτήσεις των υπόλοιπων ενότη-

των, έτσι ώστε να μη δημιουργείται σύγχυση και να είναι εφικτός ο έλεγχος της εγκυρότητας και της αξιοπιστίας για κάθε ενότητα ξεχωριστά. Είναι σαφές ότι οι ερωτήσεις μιας ενότητας ενός ερωτηματολογίου δεν πρέπει να αναμειγνύονται με τις ερωτήσεις οποιασδήποτε άλλης ενότητας του ερωτηματολογίου. Για παράδειγμα, δεν πρέπει σε μια ενότητα να περιλαμβάνονται ταυτόχρονα ερωτήσεις που αφορούν σε δημογραφικά χαρακτηριστικά και ερωτήσεις που αφορούν στην εκτίμηση της ποιότητας ζωής.

5. ΣΥΝΟΨΗ

Επισημαίνεται ότι αρχικά ελέγχεται η εγκυρότητα ενός ερωτηματολογίου και εφ' όσον εμφανίσει αποδεκτή εγκυρότητα, στη συνέχεια ελέγχεται και η αξιοπιστία του. Είναι σαφές ότι αρχικά πρέπει να ελέγχεται εάν ένα ερωτηματολόγιο μετρά την έννοια ή, αλλιώς, τη μεταβλητή την οποία διατείνεται ότι μετρά, καθώς εάν, π.χ., ένα ερωτηματολόγιο διατείνεται ότι μετρά την ποιότητα ζωής, αλλά εμφανίσει μη αποδεκτή εγκυρότητα, τότε σημαίνει ότι δεν είναι

έγκυρο για τη μέτρηση της ποιότητας ζωής και ότι πρέπει να τροποποιηθεί κατάλληλα. Με την προϋπόθεση ότι ένα ερωτηματολόγιο μετρά την έννοια την οποία διατείνεται ότι μετρά, ότι είναι δηλαδή έγκυρο, το ερώτημα πλέον είναι εάν μετρά την έννοια αυτή με σταθερότητα ή, αλλιώς, με συνέπεια, εάν δηλαδή είναι και αξιόπιστο. Η εμφάνιση αποδεκτής εγκυρότητας και αξιοπιστίας ενός ερωτηματολογίου σε μια μελέτη το καθιστά πλέον αποδεκτό για χρήση στη συγκεκριμένη μελέτη.

Επί πλέον, ο έλεγχος της εγκυρότητας και της αξιοπιστίας ενός ερωτηματολογίου πρέπει να υφίσταται σε κάθε μελέτη, ανεξάρτητα εάν το ερωτηματολόγιο έχει εμφανίσει αποδεκτή εγκυρότητα και αξιοπιστία σε προγενέστερες μελέτες. Η εγκυρότητα και η αξιοπιστία ενός ερωτηματολογίου δεν αποτελούν εγγενή χαρακτηριστικά του, καθώς μεταβάλλονται μεταξύ των διαφόρων μελετών. Όσο αυξάνει, πάντως, ο αριθμός των μελετών στις οποίες ένα ερωτηματολόγιο εμφανίζει αποδεκτή εγκυρότητα και αξιοπιστία, τόσο ενισχύεται η πιθανότητα να εμφανίσει αποδεκτή εγκυρότητα και αξιοπιστία και σε μια μετέπειτα μελέτη.

ABSTRACT

Validity and reliability of questionnaires in epidemiological studies

P. GALANIS

Center for Health Services Management and Evaluation, Department of Nursing, National and Kapodistrian University of Athens, Athens, Greece

Archives of Hellenic Medicine 2013, 30(1):97–110

Questionnaires in epidemiological studies must have high validity and reliability. A preliminary pilot study with a few participants (roughly 30–50) should always be conducted in order to estimate the validity and the reliability of a questionnaire and to correct as far as possible any errors and omissions detected in this way. The validity of a questionnaire is the extent to which the questionnaire measures the study concept or, otherwise, the variate that it is supposed to measure. Increase in validity of a questionnaire is associated with a decrease in systematic error. Evaluation of the validity of a questionnaire includes evaluation of content validity, construct validity, criterion validity and face validity. Content validity is the extent to which a questionnaire covers the different dimensions of the concept that is being measured. Construct validity is the extent to which a questionnaire reflects the real theoretical meaning of the concept that it is supposed to measure. Estimation of construct validity can be performed by factor analysis, convergent validity, discriminant validity and the known groups' method. Factor analysis is a complex mathematical procedure, comprising explanatory factor analysis and confirmatory factor analysis. Criterion validity is about the use of a criterion, i.e., a questionnaire with proved validity and reliability, as a reference method or "gold standard" for estimation of the validity of a new questionnaire; the new and the proven questionnaires must measure the same concept. The reliability or precision of a questionnaire concerns the degree of stability or consistency with which the questionnaire measures the concept that it is supposed to measure. Increase in the reliability of a questionnaire is associated with a decrease in random error. The estimation of reliability of a questionnaire includes test-retest estimation, alternative forms reliability, split-half reliability, internal consistency reliability and inter-observer reliability.

Key words: Factor analysis, Questionnaire, Reliability, Validity

Βιβλιογραφία

1. OPPENHEIM AN. *Questionnaire design, interviewing and attitude measurement*. Continuum International Publishing Group, London, 2000
2. GILLHAM B. *Developing a questionnaire*. 2nd ed. Continuum International Publishing Group, London, 2007
3. SARIS WE, GALLHOFER IN. *Design, evaluation and analysis of questionnaires for survey research*. John Wiley & Sons, New Jersey, 2007
4. GREENHALGH T. *How to read a paper. The basics of evidence-based medicine*. 4th ed. John Wiley & Sons, New Jersey, 2010:177–187
5. BJORNER JB, OLSEN J. Questionnaires in epidemiology. In: Olsen J, Saracci R, Trichopoulos D (eds) *Teaching epidemiology. A guide for teachers in epidemiology, public health and clinical medicine*. 3rd ed. Oxford University Press, Oxford, 2010:93–104
6. ΓΑΛΑΝΗΣ Π. Χρησιμοποιώντας το κατάλληλο ερωτηματολόγιο στις επιδημιολογικές μελέτες. *Αρχ Ελλ Ιατρ* (υπό δημοσίευση)
7. HIGGINS PA, STRAUB AJ. Understanding the error of our ways: Mapping the concepts of validity and reliability. *Nurs Outlook* 2006, 54:23–29
8. ΓΑΛΑΝΗΣ ΠΑ, ΣΠΑΡΟΣ ΛΔ. *Εγχειρίδιο Επιδημιολογίας*. Ιατρικές Εκδόσεις ΒΗΤΑ, Αθήνα, 2010
9. ΓΑΛΑΝΗΣ ΠΑ, ΣΠΑΡΟΣ ΛΔ. *Κλινική και επιδημιολογική έρευνα. Βασικές έννοιες*. Ιατρικές Εκδόσεις ΒΗΤΑ, Αθήνα, 2012
10. BLAND JM, ALTMAN DG. Statistics notes: Validating scales and indexes. *Br Med J* 2002, 324:606–607
11. GOODWIN LD. Changing conceptions of measurement validity. *J Nurs Educ* 1997, 36:102–107
12. LYNN MR. Determination and quantification of content validity. *Nurs Res* 1986, 35:382–385
13. DRÖGE C. How valid are measurements? *Decision Line* 1996, 27:10–12
14. DRÖGE C. Assessments of validity. *Decision Line* 1997, 28:10–12
15. BOYNTON PM. Administering, analysing, and reporting your questionnaire. *Br Med J* 2004, 328:1372–1375
16. ΜΠΕΛΛΑΛΗ Θ. Κριτήρια και διαδικασία αξιολόγησης των ποιοτικών ερευνών στο χώρο της υγείας. *Αρχ Ελλ Ιατρ* 2006, 23:298–307
17. ΡΑΦΤΟΠΟΥΛΟΣ Β, ΘΕΟΔΟΣΟΠΟΥΛΟΥ Θ. Μεθοδολογία στάθμισης μιας κλίμακας. *Αρχ Ελλ Ιατρ* 2002, 19:577–589
18. ΟΥΖΟΥΝΗ Χ, ΝΑΚΑΚΗΣ Κ. Η αξιοπιστία και η εγκυρότητα των εργαλείων μέτρησης σε ποσοτικές μελέτες. *Νοσηλευτική* 2011, 50:231–239
19. GIBBON B. Validity and reliability of assessment tools. *Nurs Res* 1995, 2:48–55
20. LITWIN MS. *How to measure survey reliability and validity*. Sage Publications, London, 1995
21. SUSKIE L. *Questionnaire survey research: What works*. Association for Institutional Research, Tallahassee, 1992
22. STREINER DL, NORMAN GR. *Health measurement scales. A practical guide to their development and use*. 4th ed. Oxford University Press, Oxford, 2008
23. BRINK TL. *Questionnaires. Practical hints on how to avoid mistakes in design and interpretation*. Heuristic Books, Chesterfield, 2004
24. RUNGTUSANATHAM M. Let's not overlook content validity. *Decision Line* 1998, 29:10–13
25. LAWSHE CH. A quantitative approach to content validity. *Personnel Psychology* 1975, 28:563–575
26. REYMONT R, JORESKOG KG. *Applied factor analysis in the natural sciences*. New York, Cambridge University Press, 1993
27. BRYANT FB, YARNOLD PR. Principal-components analysis and exploratory and confirmatory factor analysis. In: Grimm LG, Yarnold PR (eds) *Reading and understanding multivariate analysis*. American Psychological Association Books, Washington, 1995:99–136
28. WARE JE, SNOW KK, KOSINSKI M, GANDEK B. *The SF-36 health survey manual and interpretation guide*. The Health Institute, New England Medical Center, Massachusetts, 1993
29. WARE JE. Measuring patients' views: The optimum outcome measure. *Br Med J* 1993, 306:1429–1430
30. ΓΑΛΑΝΗΣ Π. Στατιστικές μέθοδοι ανάλυσης δεδομένων. *Αρχ Ελλ Ιατρ* 2009, 26:699–711
31. SUSHIL S, VERMA N. Questionnaire validation made easy. *European Journal Scientific Research* 2010, 46:172–178
32. CRONBACH LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951, 16:297–334
33. BLAND JM, ALTMAN DG. Cronbach's alpha. *Br Med J* 1997, 314:572

Corresponding author:

P. Galanis, 14 Dikis street, GR-157 73 Athens, Greece
e-mail: pegalan@nurs.uoa.gr